硬 X 射线调制望远镜卫星数据分析软件的设计及实现[1]

赵海升[1,2]，聂建胤[1,2]，李小波[1,2]，葛明玉[1,2]，李承奎[1,2]，宋黎明[1,2]

（1. 中国科学院粒子天体物理重点实验室，北京 100049；2. 中国科学院高能物理研究所，北京 100049）

摘要：硬X射线调制望远镜卫星（Hard X-ray Modulation Telescope，HXMT）是一颗空间X射线观测卫星，目前已经获取了大量观测数据。作为载荷硬件和科学分析的桥梁，数据分析软件起到非常重要的作用，它与科学结果的产出息息相关。本文主要介绍HXMT数据分析软件开发的需求，分析过程及软件框架等。HXMT数据软件不仅提供了一系列工具帮助用户分析HXMT数据，同时用户可以基于HXMT框架开发自己的程序进而分析HXMT数据，更为重要的是本软件提供的底层库可以在不同卫星系统中使用。希望本工作对未来数据分析软件的设计和开发提供借鉴。

关键词： 空间X射线，卫星，数据分析，软件框架

中图分类号：P171.5　　　文献标识码：A　　　文章编号：1672-7673（2021）01

X 射线天文学的观测多以空间观测为主，自上世纪 70 年代以来，国外已经陆续发射了多颗 X 射线天文卫星。中国也在 2017 年 7 月 15 日发射了以 X 射线天体源观测为主的硬 X 射线调制望远镜卫星（Hard X-ray Modulation Telescope，HXMT）[1]。他们大多公开了卫星观测数据，同时发布数据分析软件。后续的爱因斯坦探针卫星（Einstein Probe，EP），中法太空望远镜项目（Space Variable Objects Monitor，SVOM）也已经处于工程阶段，增强型 X 射线时变与偏振天文台（enhanced X-ray Timing and Polarimetry Mission，eXTP）也处于规划阶段。这些 X 射线天文卫星的发射必将对天文卫星数据分析方法及软件带来一定的挑战。

作为载荷硬件和科学分析的桥梁，数据分析软件的开发是必不可少的。数据分析的过程是希望通过观测数据反推入射信息，而观测数据已经经过了仪器的响应，不能直接表征入射信息。这样数据分析软件的目的是产生正确的能谱、光变以及响应等。在这一基础上，用户才能利用这些信息反推天体源的入射信息，进而推断发生的物理过程。总之，用户分析软件与物理成果的产出息息相关，是保障物理成果产出的必要条件。

卫星发布的分析软件一般是专用于该卫星数据分析的软件，另外还有通用的数据分析软件。通用软件主要针对的是天文数据格式的处理（比如，ftools[2]），以及对能谱、光变处理（比如 Xspec，powspec[3]）。NASA 提供的 HEAsoft 软件包[4]集合了通用软件及一系列卫星

*基金项目：国家自然科学基金（U1938102，11503029）.
收稿日期：2021；修订日期：2021
作者简介：赵海升，男，博士. 研究方向：天文数据分析及算法. Email：zhaohs@ihep.ac.cn

的专用软件。这样卫星数据分析软件需要重点设计卫星载荷的数据处理过程，并使得产生的结果（以文件为主）可以被通用软件识别和处理。本文主要侧重介绍 HXMT 数据分析软件的需求，分析过程及底层的设计，希望为后续卫星的软件设计和开发提供借鉴。

HXMT 卫星有三个主载荷，分别为高能 X 射线望远镜（HE，晶体探测器）[5]，中能 X 射线望远镜（ME，Si-PIN 探测器）和低能 X 射线望远镜（LE，CCD 型），它们均集成了多个单体探测器，且单体之间有差异。卫星下传的数据经过地面处理[6]后分别生成各载荷的文件目录，这些文件包括轨道、姿态、事例、温度、工作状态参数文件等，数据分析软件既是对这些文件进行处理，产生物理结果（能谱，光变，响应等）的软件。本软件主要适用于 X 射线源的定点观测，这也是 HXMT 主要的观测模式，部分模块也可以适用于扫描模式，并对短时暴发（比如太阳耀斑，X 射线暴发）有一定的处理能力。

1 分析软件的需求

空间天文卫星的数据格式一般是 FITS（Flexible Image Transport System）[7]，但是 FITS 无数据分析功能，分析此格式的数据，需要借助通用的分析工具或者自己开发工具。这一特点使得用户分析软件必须给出标准的数据筛选条件以避免用户对基础数据的分析，而这种基础分析大多基于可视化界面进行分析，同时也必须给出符合通用数据分析工具的 FITS 格式数据，比如能谱的格式及关键字定义均需要符合 Xspec 的调用。

天文数据分析工具的使用已经有几十年的历史，培养了一大批用户。HXMT 数据分析工具开发早期，我们制定了运行需要遵从 HEAsoft 的形式，能够兼容 HEAsoft 工具的策略。但是需要注意的是，HEAsoft 早期一大部分开发代码是基于 Fortran，Perl 语言开发的，移植性比较差；随后一些卫星专用软件，比如 Swift 分析软件，采用了 c 语言，移植性及可读性也比较差。当然，HEAsoft 也提供了大量优秀的代码，比如 HXMT 分析软件采用了 FITS 文件读写库 cfitsio。

天文分析工具的运行单元称为任务（task），每个 task 的输入及输出如图 1 所示，其输入及输出一般是 FITS 格式数据，另外还有一个参数文件（pfile 所示），用于程序运行中的参数控制。需要指出的是 task 的运行存在两个问题：task 之间主要采用数据文件相连，频繁的读写 FITS 文件数据，处理效率大打折扣；参数文件的使用也带来另一个问题，当多个相同进程读取同一参数文件时，可能导致运行出错。集群计算（云计算）情况下，用户不得不设置多个参数文件路径（HEAsoft 下设置 PFILES 环境变量），使得不同进程读取不同的参数文件，然而对于提交同一计算机上的并行程序却无能为力。

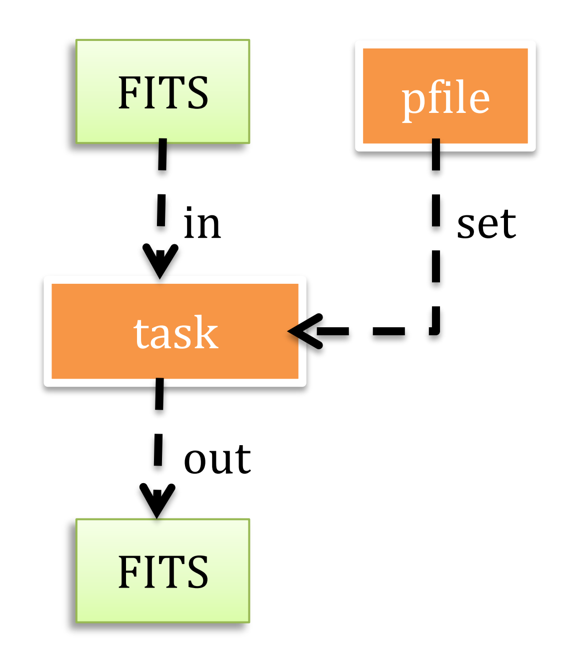


图 1 HEAsoft 下各个工具的输入及输出。

Fig. 1 The input and output of HEAsoft tools.

X 射线天文软件一般运行于 Linux 或者 Mac OS X 等操作系统上，硬盘及内存使用与数据文件大小和运行中的非线性计算有关，一般要求不大。编译系统一般可以基于 HEAsoft 下提供的 hmake 脚本完成。

另外，数据分析软件需要与标定数据库（CALDB，Calibration Database）配合使用，CALDB 主要存放与载荷、标定相关的数据文件，分析软件需要建立接口与 CALDB 相连。

这样，HXMT 需要继承 HEAsoft 的运行方式，包括数据格式，运行模式等，但是我们也希望改变 HEAsoft 的一些缺点，比如改变面向过程的编程方式，避免重复读写文件等。

2 数据分析流程

卫星下传的数据经过一系列处理，最终以单次观测数据进行组织（称之为观测号），数据分析的基本单元是观测号，HXMT 发布的是未标定、未筛选的数据，这是数据分析的输入。

数据分析软件一般遵从如图 2 的数据分析步骤，整个流程分为三步：(1)数据的标定，包括增益修正，噪声扣除，以及事例重建等方面；(2) 数据筛选，利用工程数据提取载荷正常工作的时间及源观测时间，并扣除地球遮挡等时间，然后利用事例等级及提取的好时间对标定的事例数据进行筛选，产生标定并筛选后的数据文件；(3) 高级产品提取，提取筛选后数据文件中的能道及时间信息，并根据时间、指向等信息产生能谱的响应文件，利用模型或者非源区域提取本底。这三个步骤完成后，用户可以基于这些文件进行能谱及时变方面的分析。

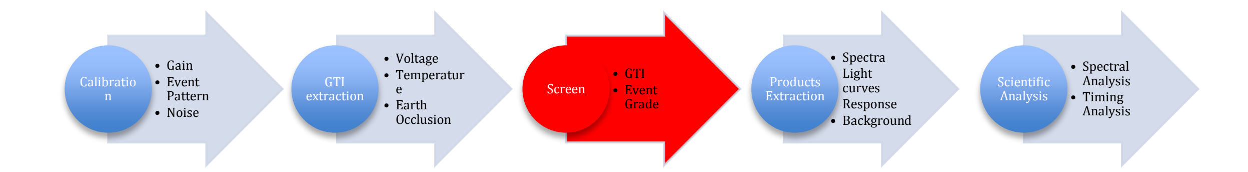


图 2 数据分析步骤，其中 GTI 为好时间段（Good Time Interval）.

Fig. 2 The data reduction flow diagram. The GTI means Good Time Interval.

HXMT 数据分析过程中还要识别和扣除一部分损坏的像素，对数据进行必要的修正，比如对光变曲线进行死时间修正及指向修正等，并计算一些重要的参量，特别是在能谱文件中曝光时间的计算是非常重要的，这是因为能谱拟合中它作为参数用于计算源的流量。

相比于其它卫星，HXMT 的三个载荷均采用了多单体探测器的集成方案增大接收面积。但是由于制造工艺的差异，以及在轨探测器温度的差异，这些探测器往往不能看为全同的。这些差异体现在数据上主要表现在好时间的不同以及增益的不同，而后者可通过增益修正改变。好时间的不同对探测器能谱（以及响应）和光变的合并带来麻烦。对于能谱，不同能道上的计数可以直接合并，但是总曝光时间并非是单体曝光的累加，而应该是按单体的权重累加，比如 HXMT 的高能探测器由 18 个 NaI 组成，权重因子可以简单的设置为 1/18。响应也是按单体给出，然后各个单体的响应按权重累加。光变需要提供四列：时间（time），计数（counts），计数误差（error）以及该时间并（binsize）中的曝光比例（fracexp）。如果需要做死时间/指向修正的话，需要先按单体计算，得到修正因子及每个 binsize 的曝光比例，然后计算单体修正后的计数及误差，最后计数累计，曝光比例按权重合并，误差按误差传递给出。

3 分析软件的设计

HXMT 卫星为天文卫星，数据格式遵从天文数据格式 FITS，分析遵从天文数据分析的特点，并可以兼容 HEAsoft 工具。

HXMT 有三个主要载荷，它们采集的均是光子事例，观测限制条件也比较一致，需要的步骤也基本相同。如果每个载荷单独开发的话，将使得整个项目费时费力。如果能够找到它们之间的相同点和不同点，并设计一种框架进行逐层次开发，这无疑有巨大的优势。当然，这三个载荷的不同设计，也决定了数据处理的不同，比如 ME 和 LE 有事例重建，而 HE 无需重建；即使是同一载荷，不同单体的差异也需要考虑。

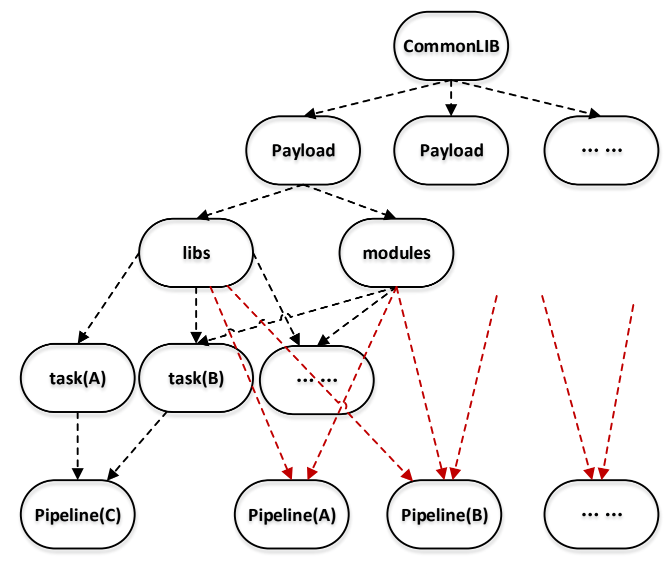


图 3 HXMT 分析软件框架。

Fig. 3 The software architecture of HXMT

HXMT 分析软件的框架如图 3 所示。架构是分层设计的，整个软件系统公共的功能放到第一层（公共层，图 3 中 CommonLIB 层），这些功能包括 task 的输入输出控制，CALDB 接口，FITS 文件的通用操作，坏像素识别类，以及与时间、能谱、光变相关的功能等；后续是以载荷进行划分层次结构，越靠后各个层将逐步考虑模块的具体实现功能（载荷层），并将本层的公共部分（图 3 中 payload，及 libs）和特殊部分（图 3 中 modules）分开，公共部分包括载荷事例、视场、探测器的定义等，特殊部分包括增益修正、死时间修正等功能类；最后一层或者多层（过程层，图 3 中 task 及 pipeline 层）实现软件的流程功能。

框架的开发采用面向对象的设计语言 C++，采用面向对象的编程形式开发，这是本软件与以前软件开发的不同之一。采用面向对象的编程的一个重要考虑是这些库（底层算法、类和接口等）将比较容易的被本软件以及被其它软件（比如用户贡献的软件）所调用。这种设计提高了软件开发质量并降低了软件开发的工作量。比如，当开发和测试载荷层时，我们会不断的使用公共层，进而加强对其的开发测试；整个系统的逻辑比较清晰，避免重复开发，这将极大降低软件代码。

设计中，HXMT 重点考虑了事例流的非线性操作及多单体探测器的处理问题。

HXMT 的载荷差异比较大，这使得分析方法迥异，数据处理框架需要解决这个问题。HXMT 通过对象与流程分开操作的方式解决此问题，特别是将流程放大以满足差异化。设计中并没有将软件的流程看为类的一部分，而是将流程看为主结构，在流程内对类进行实例化。这样的好处是灵活，即流程可以变化，比如多个流程可以整合为一个主流程。这对 HXMT 三种载荷的差异化及非线性操作非常有利，比如高能载荷在事例流上存在尖峰（毛刺事例），软件必须积累一定的数据才能识别这种尖峰并扣除，在接下来的分析中则无需考虑此因素，因此

将这部分特殊功能放在某一流程中实现，而不将其看为对象供其它分析步骤调用。

图 3 中 pipeline 工具可以通过调用多个 task 完成分析，也可以基于公共层和载荷层直接编写代码实现。用户可以不必关注这种多单体探测结构，而只需关注载荷公共的运行时间即可；用户可以不使用参数文件，进而避免并行计算存在的问题；不必将流程分为多个子流程，而使得每个子流程开发为一个工具，仅提供一个主流程即可，这样可以弥补了频繁读取 FITS 文件的过程；用户也无需在编译上依赖 HEAsoft，可以开发为独立软件。特别说明的是，这种方法对空间环境、暴发现象的研究特别有帮助。

另外针对多探测器结构，HXMT 分析软件将各个单体单独处理，比如产生每个单体的能谱，最后将能谱合并。HXMT 采用多个好时间 extension 描述整个载荷的时间信息，很多时候，这种设计造成用户比较困惑。单体的分开处理方案以及时间描述方案会造成处理效率的下降，但是这种设计对多单体却是适用的。

软件提供的公共层不仅供 HXMT 的三个载荷使用，也是可以移植到其它卫星平台中的，比如 EP 卫星的 FXT（Followed X-ray Telescope，后随 X 射线望远镜）载荷的数据分析软件直接使用了 HXMT 的部分接口及算法等。另外软件的设计框架也用于 FXT 数据分析软件的开发。软件框架的层次性结构立足于全面的思考问题，着眼于提高软件开发质量和降低软件工作量，以系统逻辑性代替重复的软件开发，以脑力代替体力。另外，框架也充分考虑到载荷的差异性，它的载荷层突出载荷特点，而过程层则考虑流程，特别是非线性操作，这种设计可以将更多载荷囊括进来。

4 HXMT 数据分析软件的实现

HXMT 数据分析软件包括 HE、ME 和 LE 数据处理软件，每个载荷按分析流程分为标定相关模块，筛选相关模块及提取高级数据产品相关模块。图 4 是 HXMT 数据分析软件的各个 task 及分析流程。这些 task 与天文通用工具 ftools 在运行方式上是一样的，其输入及输出均遵从 HEAsoft 特点。

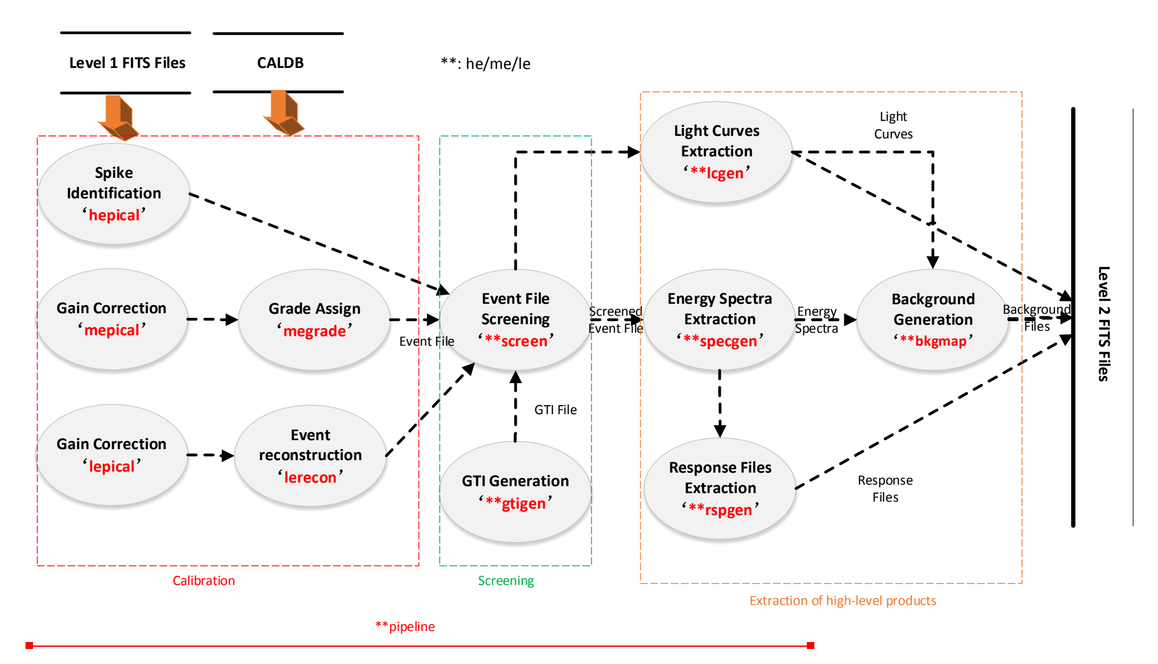


图 4 HXMT 数据分析软件的各个模块以及数据分析流程。

Fig. 4 The flow diagram and tasks of the HXMT Data Analysis Software.

HXMT 分析软件适用于定点观测，对于爆发数据的分析（主要针对 HE 载荷数据）不是太适用，比如，针对短时间观测，软件因本底难以估计进而将这些时间舍弃。其它一些与观测无关的分析，比如空间粒子，毛刺寻找等，分析软件对这些几乎无能无力，因为它提供的工具有非常强的目的性，通用性就变得很差，这也是每个卫星分析软件都存在的问题。事实上，数据分析软件不仅仅为用户提供了分析 HXMT 数据的工具，用户还可以基于 HXMT 分析软件提供的接口、算法及类等编写自己的程序。HXMT 分析软件提供的库包括：（1）与其数据格式相关的类和特定格式文件类，比如事例、电压、温度等文件的读写类，能谱，光变类等；（2）与输入输出相关类，比如运行参数的获取，日志信息的输出等；（3）与标定数据库相关类；（4）一些特定的接口类，比如死时间、视场、探测器单体描述类等；（5）特定功能类，比如对公共文件（轨道、姿态等）操作类，与时间相关类等。这些类基本满足 FITS 文件的读写，特别是事例文件的读写，以及对事例的过滤，产生高级数据产品等。

HXMT 分析软件还提供了各个载荷的 pipeline 处理模块，这些模块不是简单的将各个 task 串联起来执行，而是大量的调用软件框架所提供的接口、算法及功能类，使用一个流程将标定、筛选及提取高级数据产品整合，比如获取一个光子事例，将其进行增益修正后，判断光子信息是否符合所选择的条件，如果符合则记录光子的时间及能道信息，最后分别产生光变及能谱等。这种设计将多个流程整合为一个流程，避免了 task 之间频繁的读写 FITS 数据。用户可以借鉴此种方案开发自己的程序，同时摆脱对参数文件等的依赖。

5 总结

空间 X 射线天文卫星的数据分析软件设计大多参考 HEAsoft 软件包，并在能谱、时变分

析上基于 HEAsoft 提供的通用分析工具进行数据分析。用户也大多习惯了这种软件的运行方式及数据处理方式；天文卫星大多采用 FITS 数据格式，而大多数通用软件也以 FITS 格式的处理为主。这使得卫星专用的数据分析软件需要遵从天文数据格式，遵从天文数据分析的特点，并可以兼容 HEAsoft。HXMT 数据分析软件同样遵循了这种设计，使用上，HXMT 提供的工具与 HEAsoft 提供的工具在运行方式上并无差别，在流程上与多数专用卫星的流程保持一致。

但是 HXMT 在设计上也有其独特之处，其底层软件框架采用层次化结构。公共层提供的基础库可以为三个载荷服务，载荷层则针对载荷处理的公共和专用部分的数据处理，过程层则主要完成数据处理的流程，特别说明的是，HXMT 提供的公共层也可以应用于其它卫星的软件开发中。该框架采用面向对象的设计语言 C++开发，有很好的扩展性和继承性。HXMT 数据分析软件不仅仅提供了一系列工具用于 HXMT 数据分析，还提供了一套底层库供用户使用以开发更高效的程序代码。总之，采用面向对象的设计模式，不同卫星间公共层的移植性以及 HXMT 接口、算法和类的可调用性是 HXMT 分析软件设计的亮点，相比较而言，HEAsoft 采用面向过程的设计模式和语言，底层库比较难以调用，用户多以直接调用工具（ftools）为主。

目前 HXMT 数据分析软件已经得到国内外用户的广泛使用，其代码及使用说明可通过官网获得[8]。用户基于此已经发布大量文章[9]。框架及部分公共接口也已经用于 EP 的 FXT 数据软件开发，效果还比较不错。

参考文献

[1] Zhang, S.N., Li, T.P., Lu, F.J., et al., 2020. Overview to the hard X-ray modulation telescope (Insight-HXMT) satellite. Sci. China, Phys. Mech. Astron. 63 (4), 249502. https://doi.org/10.1007/s11433-019-1432-6.

[2] ftools: A General Package of Software to Manipulate FITS Files [OL]. [2016-05-11]. http://heasarc.gsfc.nasa.gov/ftools/.

[3] Xronos: A Timing Analysis Software Package [OL]. [2009-08-21]. http://heasarc.gsfc.nasa.gov/docs/xanadu/xronos/xronos.html

[4] NASA's HEASARC Software Packages [OL]. [2014-03-11]. http://heasarc.gsfc.nasa.gov/docs/software.html

[5] Liu, C.Z., Zhang, Y.F., Li, X.F., et al., 2020. The high energy X-ray telescope (HE) onboard the Insight-HXMT astronomy satellite. Sci. China, Phys. Mech. Astron. 63 (4), 249503. arXiv:1910.04955.

[6] 赵海升, 葛明玉, 李正恒, 聂建胤, 宋黎明. 一种天文卫星数据预处理方法[J]. 天文研究

与技术，2017，14(3)：376-381.

Zhao Haisheng, Ge Mingyu, Li Zhengheng, Nie Jianyin, Song Liming. A Method of Data Preprocessing for Astronomical Satellite[J]. Astronomical Research and Technology, 2017, 14(3)：376-381.

[7] Wells D C, Greisen E W, Harten R H. FITS: A Flexible Image Transport System [R]. AAS, 1981, 44:363-370.

[8] HXMT web, http://hxmtweb.ihep.ac.cn/software.jhtml.

[9] Insight-HXMT-publications,
https://ui.adsabs.harvard.edu/public-libraries/R_dGKc_6RSyswz9mVTsUuA.

Design and Implementation of data analysis software of Insight-HXMT satellite

Zhao Haisheng, Nie Jianyin, Li Xiaobo, Ge Mingyu, Li Chenkui, Song Liming

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract：The Hard X-ray Modulation Telescope satellite (Insight-HXMT) is a X-ray astronomy satellite, and has accumulated a large of data sets since its launch. As a bridge of hardware of payload and scientific results, the HXMT data analysis software plays an important role in obtaining scientific achievements. In this work, we will introduce the software requirement, data processing and software architecture. The HXMT data analysis software not only has provided a series of tools to help users to do data analysis of HXMT, but also provided libraries, interfaces and algorithms for user to establish a coordinated data analysis. Thus, we hope this work can provide reference for the future designation and development of data analysis software of satellite. More importantly, the common libraries and architecture can be used to future satellite software.

Keywords：Space X-ray, Satellite, data reduction, software architecture